A CLASSIFICATION OF BINARY SYSTEMATIC CODES OF SMALL DEFECT

ALBERTO RAVAGNANI[1]

Department of Mathematics, University of Neuchâtel
Rue Emile-Argand 11, CH-2000 Neuchâtel (Switzerland)

Abstract. In this paper, after having classified all the binary MDS codes, we provide and discuss two possible classifications of (non-linear) binary systematic AMDS codes, proving that they surprisingly coincide. Finally, we explicitly compute the undetected error probabilities associated to such codes.

1. Introduction

Let q be a prime power and let $\mathbb{F}_q$ denote the finite field with q elements. A (non-linear) **code** of lenght $n \in \mathbb{N}_{\geq 1}$ over the field $\mathbb{F}_q$ is simply a non-empty subset $C \subseteq \mathbb{F}_q^n$. This definition is the more general that could be given. We omit the adjective *non-linear* for the rest of the paper. The elements of C are called **codewords**, or simply **words**. We say **vector** when we generically refer to an element of the vector space $\mathbb{F}_q^n$. A code $C \subseteq \mathbb{F}_q^n$ is said to be **linear** if it is a vector subspace of $\mathbb{F}_q^n$. Linear codes can be deeply studied by using linear algebra techniques (see [11], Chapter 1, for an introduction on linear codes). Define the so-called **Hamming distance** on $\mathbb{F}_q^n$ by $d : \mathbb{F}_q^n \times \mathbb{F}_q^n \to \mathbb{R}$ with

$$d(v, w) := |\{1 \leq i \leq n : v_i \neq w_i\}|,$$

where $v = (v_1, ..., v_n)$ and $w = (w_1, ..., w_n)$. The **weight** of a vector $v \in \mathbb{F}_q^n$ is the integer $\mathrm{wt}(v) := |\{1 \leq i \leq n : v_i \neq 0\}|$, where $v = (v_1, ..., v_n)$. Let $C \subseteq \mathbb{F}_q^n$ be a code containing the zero codeword $0_n := (0, ..., 0)$. For any $i \in \mathbb{N}$ such that $0 \leq i \leq n$ we will denote by $W_i(C)$ the number of the codewords of C having weight exactly i. The collection $\{W_i(C)\}_{0 \leq i \leq n}$ is said to be the **weight distribution** of C. The $W_i(C)$'s are called **weights**. The **minimum distance** $d(C)$ of a given code $C \subseteq \mathbb{F}_q^n$ is the integer

$$d(C) := \min_{\substack{v, w \in C \\ v \neq w}} d(v, w).$$

If $C = \{v\}$ is a single-word code then we will set, by definition, $d(C) := \infty$. A code of lenght n, $|C|$ codewords and minimum distance d is said to be of **parameters** $[n, |C|, d]$. The correction-capability of a code is strictly related to its minimum distance. Indeed, a code $C \subseteq \mathbb{F}_q^n$ of minimum distance d can correct (by brute-force) $\lfloor (d-1)/2 \rfloor$ errors. The general theory of error-correcting codes is well-developed in [11] and [8].

If $C \subseteq \mathbb{F}_q^n$ is a linear code of minimum distance $d = d(C)$ and dimension k (as a vector subspace of $\mathbb{F}_q^n$) then in any case $d \leq n - k + 1$. This is the well-known Singleton bound (see [14]), which gives an upper bound for the minimum distance of a linear code of given dimension k. Write the bound in the equivalent form $k \leq n - d + 1$ and observe that it could be read as an upper bound on the dimension of a linear code of given minimum distance. In this interpretation, the result easily generalizes into a Singleton bound for non-linear codes $C \subseteq \mathbb{F}_q^n$. Indeed, if $C \subseteq \mathbb{F}_q^n$ is any code of minimum distance d then we can remove from the codewords of C any $d - 1$ components and get distinct vectors of lenght $n - d + 1$. The number of these vectors, $|C|$, cannot exceed the size of $\mathbb{F}_q^{n-d+1}$, which is q^{n-d+1}.

E-mail address: `alberto.ravagnani@unine.ch`[1].

Key words and phrases. Non-linear code, systematic code, binary code, AMDS code.

Proposition 1 (Singleton bound)**.** Let $C \subseteq \mathbb{F}_q^n$ be a code of minimum distance d. Then $|C| \leq q^{n-d+1}$.

A code $C \subseteq \mathbb{F}_q^n$ of minimum distance d and attaining the Singleton bound is said to be a **MDS** codes (MDS stands for *Maximum Distance Separable*). A code $C \subseteq \mathbb{F}_q^n$ of minimum distance d with $|C| = q^{n-d}$ words is said to be an **AMDS** code (AMDS stands for *Almost* MDS). MDS codes over arbitrary alphabets $\mathbb{F}_q = \{0, 1, ..., q-1\}$ of q symbols and containing the zero codeword have been studied in [15], providing the whole weight distribution without introducing any algebraic structure on the alphabet.

2. Relevant properties of MDS and AMDS codes

Throughout this section, n will denote a fixed positive integer and $\mathbb{F}_q$ the finite field with q elements. For any non-empty set $S \subseteq \{1, 2, ..., n\}$ of cardinality s we denote by $\pi_S : \mathbb{F}_q^n \to \mathbb{F}_q^s$ the projection on the coordinates appearing in S. The following two properties of MDS codes are easily proved.

Lemma 2. Let $C \subseteq \mathbb{F}_q^n$ be any code of minimum distance d.

(1) The code C is a MDS code if and only if for any $S \subseteq \{1, 2, ..., n\}$ of cardinality $n - d + 1$ we get $\pi_S(C) = \mathbb{F}_q^{n-d+1}$.
(2) Assume that C is a MDS code and set $k := n - d + 1$, so that C has q^k elements. There exists a function $\varphi : \mathbb{F}_q^k \to \mathbb{F}_q^{d-1}$ such that $C = \{(v, \varphi(v)) : v \in \mathbb{F}_q^k\}$. Moreover, C is a linear code if and only if φ is a linear map.

In the next lemma we briefly recall two well-known combinatoric bounds for non-linear codes.

Lemma 3. Let $C \subseteq \mathbb{F}_q^n$ be any code of minimum distance d. The following facts hold.

(1) If $|C| = q^k$ then $d \leq \frac{nq^k(q-1)}{q(q^k-1)}$ (Plotkin bound),
(2) $|C| \cdot \sum_{i=0}^{\lfloor (d-1)/2 \rfloor} \binom{n}{i}(q-1)^i \leq q^n$ (Hamming bound).

Proof. See for instance[1] [12], Theorems 1.1.45 and 1.1.47. □

As an application of the Hamming bound we obtain powerful restrictions on the size of MDS and AMDS codes.

Proposition 4. Let $C \subseteq \mathbb{F}_q^n$ be a code of minimum distance $d \geq 3$ and $|C| = q^k$ words.

(1) If C is a MDS code then $k \leq q - 1$.
(2) If C is an AMDS code then $k \leq q^2 + q - 2$.

Proof. Set $s(C) := n - d - k + 1$, so that $s(C) = 0$ if C is a MDS code and $s(C) = 1$ if C is an AMDS code. Remove from the codewords of C any $d - 3$ components. In this way we get a code D of lenght $n - d + 3$, minimum distance at least 3 and $|C| = q^k$ codewords. Applying the Hamming bound to the code D we get

$$q^{n-d+1-s(C)} [1 + (n - d + 3)(q - 1)] \leq q^{n-d+3}.$$

Straightforward computations lead to the thesis. □

Remark 5. The parameter $s(C)$ introduced in the proof of Proposition 4 is called the **Singleton defect** of the code C. Linear codes of small defects have been studied by Faldum and Willems in [6]. The lenght of certain linear AMDS codes has been deeply discussed in [1] and in [5].

3. The binary case

Let $\mathbb{F}_2 := \{0, 1\}$ be the finite field with two elements ($q = 2$). For any $r \geq 1$ we will denote by 0_r the word made of r 0's and by 1_r the word made of r 1's. Denote by n a fixed positive integer.

Definition 6. Let $u \in \mathbb{F}_2^n$ be any vector. The **translation** by u is the map $T_u : \mathbb{F}_2^n \to \mathbb{F}_2^n$ defined by $T_u(v) := u + v$. If $C \subseteq \mathbb{F}_2^n$ is a code and $u \in \mathbb{F}_2^n$ then $T_u(C) \subseteq \mathbb{F}_2^n$ denote the image of C under T_u.

[1] For interesting discussions about bounds obtained by code coverings see [4], Chapters 6 to 14.

The following lemma summarizes some obvious properties of translations over $\mathbb{F}_2$.

Lemma 7. The following facts hold.

(1) For any $u, v \in \mathbb{F}_2^n$, $T_u(v) = T_v(u)$.
(2) For any $u \in \mathbb{F}_2^n$, the map $T_u : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is bijective.
(3) For any $u \in \mathbb{F}_2^n$, $T_u \circ T_u = \mathrm{Id}_{\mathbb{F}_2^n}$, the identity of $\mathbb{F}_2^n$.
(4) For any $u, v, w \in \mathbb{F}_2^n$, $d(v, w) = d(T_u(v), T_u(w))$, d being the Hamming distance on $\mathbb{F}_2^n$.

Translations define as follows a very strong equivalence relation of binary codes (the definition is easily extended to arbitrary finite fields).

Definition 8. Codes $C, D \subseteq \mathbb{F}_2^n$ are said to be T**-equivalent** if there exists a vector $u \in \mathbb{F}_2^n$ such that $C = T_u(D)$.

Lemma 9. Any two T-equivalent codes in $\mathbb{F}_2^n$ have the same number of codewords and the same minimum distance. Every code $C \subseteq \mathbb{F}_2^n$ is T-equivalent to a code containing the word 0_n. T-equivalent codes $C, D \subseteq \mathbb{F}_2^n$ containing 0_n have the same weight distribution.

Proof. The first part and the last one trivially follow from Lemma 7. For the second part of the statement, pick out any $u \in C$ and observe that $0_n \in T_u(C)$. □

4. Classification of binary MDS codes

In this short section we classify all the binary MDS codes (i.e. MDS codes over $\mathbb{F}_2$) by using only the metric properties of the Hamming distance and translations.

Lemma 10. Let $C \subseteq \mathbb{F}_2^n$ be a MDS code of minimum distance d and such that $0_n \in C$. Set $k := n - d + 1$. If $d = 2$ then C is the parity-check code of the code $\mathbb{F}_2^k$.

Proof. By Lemma 2 there exists a function $\varphi : \mathbb{F}_q^2 \to \mathbb{F}_2$ such that $C = \{(v, \varphi(v)) : v \in \mathbb{F}_q^k\}$. We have to prove that φ is the parity-check function, i.e.

$$\varphi(v) = \begin{cases} 0 & \text{if } \operatorname{wt}(v) \equiv 0 \pmod 2 \\ 1 & \text{if } \operatorname{wt}(v) \equiv 1 \pmod 2 \end{cases}$$

for any $v \in \mathbb{F}_2^k$. Let us proceed by induction on the weight of $v \in \mathbb{F}_2^k$. Since $0_n \in C$ we get $\varphi(0_k) = 0$. Let $v \in \mathbb{F}_2^k$ be any vector of weight $0 < t \leq n$. Assume that φ is defined to be the parity-check function on the vectors of $\mathbb{F}_2^k$ of weight $t-1$. There obviously exists a vector $w \in \mathbb{F}_2^k$ of weight $t-1$ and such that $d(v, w) = 1$. Since $d = 2$ we deduce $\varphi(v) \neq \varphi(w)$. Since $\varphi(w)$ depends only on the weight of w (which is $t - 1$) we get the thesis. □

Theorem 11 (T-classification of binary MDS codes)**.** Let $C \subseteq \mathbb{F}_2^n$ be any MDS code of minimum distance d. Then C is T-equivalent to exactly one of the following MDS codes.

(1) The n-times repetition code with $d = n \geq 3$.
(2) The parity-check code of the code $\mathbb{F}_2^k$, $k = n - 1$.
(3) The code $\mathbb{F}_2^n$.

Proof. By T-equivalence we shall assume $0_n \in C$ (Lemma 9). If $d \geq 3$ then apply Proposition 4 and deduce that C has two codewords and $d = n$. Since $0_n \in C$, the other word of C must be 1_n, i.e. C is the n-times repetition code. Assume $d = 2$. By Lemma 10 we get that C is the parity-check code of $\mathbb{F}_2^k$, with $k = n - d + 1 = n - 1$. Assume $d = 1$. Since C is MDS we must have $|C| = 2^{n-1+1} = 2^n$. Since $C \subseteq \mathbb{F}_2^n$ we conclude $C = \mathbb{F}_2^n$. Finally, one can easily check that the three sets of codes listed in the statement are pairwise disjoint. □

Corollary 12. Any MDS code $C \subseteq \mathbb{F}_2^n$ is an affine subspace of $\mathbb{F}_2^n$.

Remark 13. Corollary 12 has been proved also in [7] by using an original argument on Grœbner bases and systematic codes. The proof given in this paper reveals, on the other hand, the purely metric nature of the classification.

5. Classifying binary systematic AMDS codes

In this section we start studying the more complicated case of binary systematic AMDS codes. A complete classification of them is provided in the rest of the paper. Systematic codes turn out to be very useful in the applications (see [13] and [2]) and strong bounds on their parameters have been recently discovered (see [3]). Let us briefly recall some basic definitions.

Definition 14. Let n be a positive integer and q a prime power. A code $C \subseteq \mathbb{F}_q^n$ is said to be **systematic** if there exists a function $\varphi : \mathbb{F}_q^k \to \mathbb{F}_q^{n-k}$ such that $C = \{(v, \varphi(v)) : v \in \mathbb{F}_q^k\}$. The function φ is a **systematic encoding function**. A code as in the definition has q^k codewords.

Remark 15. Lemma 2 says that any MDS code is in fact systematic.

Lemma 16. Any binary systematic code $C \subseteq \mathbb{F}_2^n$ is T-equivalent to a binary systematic code containing the zero codeword.

Proof. Choose a codeword $u \in C$. Write $u = (u_k, u_{n-k}) \in \mathbb{F}_2^k \times \mathbb{F}_2^{n-k}$. Apply the results of Lemma 7 to get that $T_u(C) = \{(v, \tilde{\varphi}(v)) : v \in \mathbb{F}_2^k\}$, with $\tilde{\varphi} = T_{u_{n-k}} \circ \varphi \circ T_{u_{n-k}}$. Finally, observe that $0_n \in T_u(C)$. □

The T-equivalence notion adopted in the classification of MDS binary codes is a too restrictive criterion for the classification of AMDS codes. Hence we introduce two other natural and well-known equivalence relations of codes.

Definition 17. Codes C, D over a finite field $\mathbb{F}_q$ are said to be P-equivalent if they have the same parameters:

$$[n(C), |C|, d(C)] = [n(D), |D|, d(D)].$$

Codes C, D over $\mathbb{F}_q$ and containing the zero codeword are said to be W-equivalent if they have the same lenght and the same weight distribution (see Section 1).

Remark 18. The P- and the W- equivalence are in fact equivalence relations. If $C, D \subseteq \mathbb{F}_2^n$ are T-equivalent then they are also P-equivalent. Hence, by Lemma 16, any binary systematic code is P-equivalent to a binary systematic code containing the codeword 0_n. Finally, we observe that if C and D are W-equivalent then $|C| = |D| = \sum_{i=0}^n W_i(C) = \sum_{i=0}^n W_i(D)$. If C and D are linear W-equivalent codes then they are also P-equivalent (the minimum distance agrees with the minimum weight in this case). This fact is not true in general for non-linear codes.

In the following section we explicitly describe the binary systematic AMDS codes containing the zero codeword of minimum distance $d = 1, 2$ and then P-classify all the binary systematic AMDS codes. In the next and more computational Section 7 we will derive also the complete W-classification of binary systematic AMDS codes of minimum distance $d \geq 3$.

Notation 19. For the sake of brevity, we will denote by $\mathcal{C}(n, d)$ the set of binary systematic codes of lenght $n \geq 2$ and minimum distance d and by $\mathcal{C}_0(n, d)$ the set of codes belonging to $\mathcal{C}(n, d)$ and containing the zero codeword.

6. P-classification of AMDS binary systematic codes

We start the P-classification of binary systematic AMDS codes providing an explicit characterization of the set $\mathcal{C}_0(n, 1)$.

Proposition 20. A code $C \subseteq \mathbb{F}_2^n$ of minimum distance $d = 1$ and 2^k codewords ($k \geq 1$) belongs to $\mathcal{C}_0(n, 1)$ if and only if there exists a function $\varphi : \mathbb{F}_2^k \to \mathbb{F}_2$ with the following properties:

(1) $\varphi(0_k) = 0$,
(2) φ is not the parity-check function,
(3) $C = \{(v, \varphi(v)) : v \in \mathbb{F}_2^k\}$.

As a consequence, for any $n \geq 2$, $|\mathcal{C}_0(n, 1)| = 2^{2^{n-1}-1} - 1$.

Proof. Assume that C is systematic, AMDS and contains 0_n. In the notations of Definition 14 we have $\varphi(0_k) = 0$. If φ is the parity-check function then consider two vectors $v, w \in \mathbb{F}_2^k$ such that $d(v, w) = 1$. In other words, w is obtained from v by changing one of the components. Hence $\mathrm{wt}(v) \not\equiv \mathrm{wt}(w) \pmod 2$ and $d((v, \varphi(v)), (w, \varphi(w))) = 2$. This proves that the minimum distance of C is two, a contradiction. Now assume that $\varphi : \mathbb{F}_2^k \to \mathbb{F}_2$ is not the parity-check function and $\varphi(0_k) = 0$. We have to prove that the code $C := \{(v, \varphi(v)) : v \in \mathbb{F}_2^k\}$ belongs to $\mathcal{C}_0(k+1, 1)$. The fact that C is systematic and contains the zero codeword is obvious. Assume that C is not AMDS. Since its minimum distance $d(C)$ trivially satisfies $1 \leq d(C) \leq 2$, it has 2^k elements and lenght $k+1$, this means that C is a MDS code. But this is absurd by the classification proved in Theorem 11 (φ is not the parity-check function here). □

The following result collects the study of $\mathcal{C}_0(n, 2)$ and $\mathcal{C}_0(d+1, d)$.

Proposition 21. The following facts hold.

(1) A code $C \subseteq \mathbb{F}_2^n$ of minimum distance $d = 2$ and 2^k elements ($k \geq 2$) belongs to $\mathcal{C}_0(n, 2)$ if and only if there exists a function $\varphi : \mathbb{F}_2^k \to \mathbb{F}_2^2$ with the following properties:
   (a) $\varphi(0_k) = 0$,
   (b) $\varphi(v) \neq \varphi(w)$ for any $v, w \in \mathbb{F}_2^k$ such that $d(v, w) = 1$,
   (c) $C = \{(v, \varphi(v)) : v \in \mathbb{F}_2^k\}$.

(2) A subset $C \subseteq \mathbb{F}_2^n$ is an AMDS systematic code with two elements containing 0_n if and only if it is of the form $\{0_n, (1, v)\}$ with $n \geq 2$ and $\mathrm{wt}(v) = n - 2$.

Proof. If $C \in \mathcal{C}_0(n, 2)$ then the encoding function φ of Definition 14 must obviously satisfy the properties in the statement. On the other hand, assume that $\varphi : \mathbb{F}_2^k \to \mathbb{F}_2^2$ satisfies (a), (b), (c). By (a) the systematic code $C := \{(v, \varphi(v)) : v \in \mathbb{F}_2^k\}$ contains 0_n. By (b) the code C has not minimum distance one. By the Singleton bound we have $d(C) \in \{2, 3\}$. If C has minimum distance three then it is a MDS code. On the other hand, Theorem 12 states that a binary MDS code of parameters $[k+2, 2^k, 3]$ cannot exist ($k \geq 2$). Hence $d(C) = 2$ and C is an AMDS code. The second part of the claim is obvious. □

Now we focus on the P-classification of codes in $\mathcal{C}(n, d)$ with $d \geq 3$. The following result shows that the size of any AMDS binary code with correction capability at least one has to be very small.

Lemma 22. Let $C \in \mathcal{C}(n, d)$, with $d \geq 3$. Then $k := n - d \in \{1, 2, 3, 4\}$. Moreover, if $k \in \{2, 3, 4\}$ then $d \in \{3, 4\}$.

Proof. Proposition 4 gives $k \in \{1, 2, 3, 4\}$. If $k \in \{2, 3\}$ the Plotkin bound (Lemma 3) implies $d \in \{3, 4\}$. If $k = 4$ then the same bound gives $d \in \{3, 4, 5\}$. The case $k = 4$ and $d = 5$ is excluded by the Hamming bound (Lemma 3). □

Lemma 23. Let n be a positive integer and let $v, w \in \mathbb{F}_2^n$. Then $d(v, w) = \mathrm{wt}(v) + \mathrm{wt}(v) - 2v \cdot w$, where $v \cdot w = |\{1 \leq i \leq n : v_i = w_i = 1\}|$. In particular, the integer $\mathrm{wt}(v) - \mathrm{wt}(w)$ is odd if and only if $d(v, w)$ is odd.

Proof. Working on $\mathbb{F}_2$, we have $d(v, w) = \mathrm{wt}(v + w)$. □

Theorem 24 (P-classification)**.** Any binary systematic AMDS code belongs to exactly one of the following non-empty and pairwise disjoint sets.

(1) $\mathcal{C}(n, 1)$ with $n \geq 3$,
(2) $\mathcal{C}(n, 2)$ with $n \geq 4$,
(3) $\mathcal{C}(d+1, d)$ with $d \geq 1$,
(4) $\mathcal{C}(5, 3)$,
(5) $\mathcal{C}(6, 4)$,
(6) $\mathcal{C}(6, 3)$,
(7) $\mathcal{C}(7, 4)$,
(8) $\mathcal{C}(7, 3)$,
(9) $\mathcal{C}(8, 4)$.

Proof. The fact that the listed sets are pairwise disjoint is easily checked. By Remark 18, Proposition 20 and Proposition 21 P-describe $\mathcal{C}(n, 1)$ and $\mathcal{C}(n, 2)$, providing the first three sets in the statement. By Lemma 22

the other possibly non-empty sets $\mathcal{C}(n,d)$ are only the other six of the list. Hence it is enough to show that each of the sets in the statement contains at least one code.

The set $\mathcal{C}(n,1)$ with $n \geq 3$ is non-empty by Lemma 20. Define $\varphi : \mathbb{F}_2^k \to \mathbb{F}_2^2$ by $\varphi := (p, \varphi')$, where p is the parity-check function and φ' is arbitrary with $\varphi'(0_k) = 0$. By Lemma 21 this shows that $\mathcal{C}(n,2)$ with $n \geq 4$ is non-empty. For $\mathcal{C}(d+1,d)$ with $d \geq 1$ use the second part of Lemma 21. For any $k \geq 2$ and $d \geq 3$ odd, the parity-check code of a code in $\mathcal{C}(k+d,d)$ is a code in $\mathcal{C}(k+d+1,d+1)$ (Lemma 23). Hence it is enough to show that $\mathcal{C}(5,3)$, $\mathcal{C}(6,3)$ and $\mathcal{C}(7,3)$ are non-empty. For each tern $[5,2^2,3]$, $[6,2^3,3]$ and $[7,2^4,3]$ we will provide in Table 1 a generator matrix of a binary linear systematic code having these parameters.

Table 1. Examples of linear codes in $\mathcal{C}(5,3)$, $\mathcal{C}(6,3)$ and $\mathcal{C}(7,3)$ given by their generator matrices.

$[n, 2^k, d]$ $\to$	$[5,2^2,3]$	$[6,2^3,3]$	$[7,2^4,3]$
Matrix $\to$	$\begin{bmatrix} 0&1&0&1&1 \\ 1&0&1&0&1 \end{bmatrix}$	$\begin{bmatrix} 0&0&1&0&1&1 \\ 0&1&0&1&0&1 \\ 1&0&0&1&1&0 \end{bmatrix}$	$\begin{bmatrix} 1&0&0&0&1&1&0 \\ 0&1&0&0&1&0&1 \\ 0&0&1&0&0&1&1 \\ 0&0&0&1&1&1&1 \end{bmatrix}$

The three examples conclude the proof. □

Remark 25. We point out that the code of parameters $[7,2^4,3]$ in the proof of Theorem 24 is the Hamming code $H(q=2, r=3)$ (see [11], pag. 23).

7. W-classification of AMDS binary systematic codes

We conclude the paper providing the whole W-classification of binary systematic AMDS codes containing the zero codeword, having minimum distance at least three and at least four codewords (the only relevant cases for applications). We will show, in particular, that the P- and the W- classification surprisingly coincide. In general, this fact is very far from the truth. Our proof carries also a very simple computational part written by using the Computer Algebra System `MAGMA` (see the home page `http://magma.maths.usyd.edu.au/magma/`). Since we assume $d \geq 3$ and $k > 1$, by Lemma 22 it is enough to study the weight distributions of codes belonging to the non-empty sets $\mathcal{C}_0(5,3)$, $\mathcal{C}_0(6,4)$, $\mathcal{C}_0(6,3)$, $\mathcal{C}_0(7,4)$, $\mathcal{C}_0(7,3)$ and $\mathcal{C}_0(8,4)$. We will treat the first four sets by using a computer, because the computations take only a few seconds on a common laptop (we used a processor Intel Atom N570, CPU 1.66 GHz and RAM-memory 1.0 GB in writing the paper).

Lemma 26. The weight distribution of any code $C \in \mathcal{C}_0(n,d)$, with $(n,d) \in \{(5,3),(6,4),(6,3),(7,4)\}$, depends only on the pair (n,d) itself. Moreover, the possible weight distributions are exactly those listed in Table 2. Every code in $\mathcal{C}_0(n,d)$, with $(n,d) \in \{(5,3),(6,4),(6,3),(7,4)\}$ is a linear code.

Table 2. Partial W-classification of binary systematic AMDS codes.

Set $\mathcal{C}_0(n,d)$	Non-zero weights of any $C \in \mathcal{C}_0(n,d)$
$\mathcal{C}_0(5,3)$	$W_0(C) = 1$, $W_3(C) = 2$, $W_4(C) = 1$
$\mathcal{C}_0(6,4)$	$W_0(C) = 1$, $W_4(C) = 3$
$\mathcal{C}_0(6,3)$	$W_0(C) = 1$, $W_3(C) = 4$, $W_4(C) = 3$
$\mathcal{C}_0(7,4)$	$W_0(C) = 1$, $W_4(C) = 7$

Proof. The proof is computational. Here we show, as an example, a `MAGMA` program which solves the case $(n,d) = (5,3)$. The computations for the other three sets of codes are similar. The algorithm finds out all the codes in $\mathcal{C}_0(5,3)$ and computes their weight distributions. The linearity is checked by using the following elementary fact. A subset $C \subseteq \mathbb{F}_2^n$ with q^k elements is a vector subspace of $\mathbb{F}_q^n$ if and only if the matrix whose rows are the elements of C has rank k.

```
n:=5;
dist:=3;
k:=n-dist;
for i1 in [0..1] do;
for i2 in [0..1] do;
for i3 in [0..1] do;
for i4 in [0..1] do;
for i5 in [0..1] do;
for i6 in [0..1] do;
for i7 in [0..1] do;
for i8 in [0..1] do;
for i9 in [0..1] do;
L:=[0,0,0,0,0,  0,1,i1,i2,i3,
    1,0,i4,i5,i6,  1,1,i7,i8,i9];
M:=Matrix(GF(2),2^k,n,L);
d:=100;
for j1 in [1..2^k-1] do;
for j2 in [j1+1..2^k] do;
counter:=0;
for j in [1..n] do;
if M[j1][j] eq M[j2][j] then;
counter:=counter+0;
else counter:=counter+1;
end if;
end for;
if counter le d then d:=counter;
end if;
end for;
end for;
if d eq dist then;
M;
W:=[0,0,0,0,0,0,0,0,0];
for a in [1..2^k] do;
w:=0;
for p in [1..n] do;
if M[a][p] eq 1 then w:=w+1;
end if;
end for;
W[w+1]:=W[w+1]+1;
end for;
W;
if Rank(M) eq k then printf " linear
";
else printf " not linear
";
end if;
end if;
end for;
end for;
end for;
end for;
end for;
end for;
end for;
end for;
end for;
```

The computational analysis of the others sets of codes $\mathcal{C}_0(6,4)$, $\mathcal{C}_0(6,3)$ and $\mathcal{C}_0(7,4)$ leads to the thesis. □

A similar MAGMA program could be written also to study the sets $\mathcal{C}_0(7,3)$ and $\mathcal{C}_0(8,4)$. A problem could be that such a program does not finish its computations in a reasonable time on a common computer[2]. As a consequence, we are going to solve the problem theoretically.

Lemma 27. Any code $C \in \mathcal{C}_0(7,3)$ has the following weight distribution.

$$\begin{array}{r|l} W_0(C) = 1 & W_4(C) = 7 \\ W_1(C) = 0 & W_5(C) = 0 \\ W_2(C) = 0 & W_6(C) = 0 \\ W_3(C) = 7 & W_7(C) = 1 \end{array}$$

Proof. A code $C \in \mathcal{C}_0(7,3)$ has lenght 7, $2^4 = 16$ codewords and minimum distance exactly 3. Hence C is a code attaining the Hamming bound of Proposition 3. Such a code is said to be a **perfect** code (see [11], Chapter 6 and [4], Chapter 11). By [11], Theorem 37 at page 182 and the following remark, C has the same weight distribution of the well-known Hamming code $H(q=2, r=3)$ of parameters $[7, 2^4, 3]$ (see [11], pag. 23). The weight distribution of this simple linear code is easily computed.

□

Remark 28. The weight distributions of Hamming codes over arbitrary finite fields have been deeply studied in [9] and [10].

In order to conclude the analysis we briefly recall the basic notion of distance distribution of a code.

---

[2]We estimate that a program able to study $\mathcal{C}_0(7,3)$ and analogous to that in the proof of Lemma 26 takes about four years of computations on a processor Intel Atom N570, CPU 1.66 GHz and RAM-memory 1.0 GB.

Definition 29. Let $C \subseteq \mathbb{F}_q^n$ be a code over a finite field $\mathbb{F}_q$. For any $i \in \{0, 1, ..., n\}$ define $B_i(C)$ as

$$B_i(C) := \frac{1}{|C|} \, |\{(v, w) \in C^2 : d(v, w) = i\}|.$$

The collection $\{B_i(C)\}_{i=0}^n$ is called the **distance distribution** of C.

Lemma 30. A code C in $\mathcal{C}_0(8, 4)$ is the parity check-code of a code in $\mathcal{C}_0(7, 3)$. In particular, it has the following weight distribution.

$$\begin{array}{l|l|l} W_0(C) = 1 & W_3(C) = 0 & W_6(C) = 0 \\ W_1(C) = 0 & W_4(C) = 14 & W_7(C) = 0 \\ W_2(C) = 0 & W_5(C) = 0 & W_8(C) = 1 \end{array}$$

Proof. Let $C \in \mathcal{C}_0(8, 4)$. Denote by E the code obtained by removing from the codewords of C the last component. The code E is systematic, contains 0_{n-1}, has $2^{k'}$ codewords (with $k' = k = 4$) and lenght $n' = n - 1 = 7$. Moreover, $d' := d(E) \in \{4, 3\}$. If $d' = 4$ then E is a MDS code of parameters $[7, 2^4, 4]$, which contradicts the classification of Theorem 11. Hence $d' = 3$. As in the proof of Lemma 27, E is a perfect code and by [11], Theorem 37 at page 182 (and the following remark) it has the same weight and distance distribution of the Hamming code $H(q = 2, r = 3)$:

$$\begin{array}{l|l} W_0(E) = 1 & B_0(E) = 1 \\ W_1(E) = 0 & B_1(E) = 0 \\ W_2(E) = 0 & B_2(E) = 0 \\ W_3(E) = 7 & B_3(E) = 7 \\ W_4(E) = 7 & B_4(E) = 7 \\ W_5(E) = 0 & B_5(E) = 0 \\ W_6(E) = 0 & B_6(E) = 0 \\ W_7(E) = 1 & B_7(E) = 1 \end{array}$$

Obviously there exists a function $f : E \to \mathbb{F}_2$ such that $C = \{(e, f(e)) : e \in E\}$. We will show that f is the parity-check function on E, concluding the proof. Since $0_n \in C$ we must have $f(0_{n-1}) = 0$. Since $d(C) = 4$ we get $f(e) = 1$ for any $e \in E$ such that $\mathrm{wt}(e) = 3$. Let $e \in E$ be any codeword of weight four. By Lemma 23 and the distance distribution above we deduce that $d(e, e') \in \{3, 7\}$ for any $e' \in E$ of weight three. The equation $d(e, e') = 7$ with $e \in E$ of weight four fixed and $e' \in E$ unknown of weight three has at most one solution in E (namely, $e' = T_{1_n}(e)$). Since $W_3(E) = 7$ there must exist a codeword $e' \in E$ such that $\mathrm{wt}(e') = 3$ and $d(e, e') = 3$. We have already shown that $f(e') = 1$ and, by hypothesis, $d = d(C) = 4$. Hence it must be $f(e) = 0$. Since e is arbitrary we get $f(e) = 0$ for any $e \in E$ with $\mathrm{wt}(e) = 4$. Obviously $f(1_7) = 1$ at this point. $\square$

Theorem 31 (W-classification)**.** Any binary systematic AMDS code containing the zero codeword, of minimum distance at least three and at least four codewords has exactly one of the weight distributions listed in Table 3. Moreover, each of those weight distribution corresponds to a binary systematic AMDS code containing the zero codeword.

Proof. Combine Lemma 26, Lemma 27, Lemma 30 and Theorem 24. $\square$

Table 3. W-classification of relevant binary systematic AMDS codes.

$[n, 2^k, d] \rightarrow$	$[5, 2^2, 3]$	$[6, 2^2, 4]$	$[6, 2^3, 3]$	$[7, 2^3, 4]$	$[7, 2^4, 3]$	$[8, 2^4, 4]$
$W_0(C)$	1	1	1	1	1	1
$W_1(C)$	0	0	0	0	0	0
$W_2(C)$	0	0	0	0	0	0
$W_3(C)$	2	0	4	0	7	0
$W_4(C)$	1	3	3	7	7	14
$W_5(C)$	0	0	0	0	0	0
$W_6(C)$	-	0	0	0	0	0
$W_7(C)$	-	-	-	0	1	0
$W_8(C)$	-	-	-	-	-	1

8. Undetected error probability

In this last section we explicitly compute the undetected error probability associate to each of the codes described by Theorem 31. We assume we are working on a binary symmetric channel and that all the codewords have the same probability to be transmitted.

Definition 32. Let $C \subseteq \mathbb{F}_2^n$ be a binary code and let $p < 1/2$ be a real number (the probability that 0 bicomes 1, and viceversa, during the transmission). Let us enumerate the codewords of C, say $C = \{c_1, c_2, ..., c_{|C|}\}$. The **undetected error probability** associated to C and p (see [16], Section II) is the number

$$\mathrm{UEP}(p, C) := \sum_{j=1}^{|C|} \frac{1}{|C|} \sum_{i=1}^{n} N_C(j, i)\ p^i\ (1-p)^{n-i},$$

where $N_C(j, i) := |\{c \in C : d(c_j, c) = i\}|$.

Remark 33. Let $C \subseteq \mathbb{F}_2^n$ be any code and let $\{B_i(C)\}_{i=0}^n$ be its distance distribution (Definition 29). Then it is easy to see that

$$\mathrm{UEP}(p, C) := \sum_{i=1}^{n} B_i(C)\ p^i\ (1-p)^{n-i}.$$

If $C \subseteq \mathbb{F}_2^n$ is a linear code then for any codeword $c \in C$ and for any integer $1 \leq i \leq n$ the map

$$\{v \in C : d(v, c) = i\} \rightarrow \{v \in C : \mathrm{wt}(v) = i\}$$

defined by $v \mapsto v + c$ is a bijection. Hence in the notations of Definition 32 we have $N_C(j, i) = W_i(C)$ for any $1 \leq j \leq |C|$. In particular, $N_C(j, i)$ does not depend on j. It follows

$$\begin{aligned} \mathrm{UEP}(p, C) &= \sum_{j=1}^{|C|} \frac{1}{|C|} \sum_{i=1}^{n} N_C(j, i)\ p^i\ (1-p)^{n-i} \\ &= \sum_{j=1}^{|C|} \frac{1}{|C|} \sum_{i=1}^{n} W_i(C)\ p^i\ (1-p)^{n-i} \\ &= \sum_{i=1}^{n} W_i(C)\ p^i\ (1-p)^{n-i}. \end{aligned}$$

Theorem 34. Let $p < 1/2$ be a real number. The undetected error probability associated to a code $C \in \mathcal{C}_0(n, d)$ with $(n, d) \in \{(5, 3), (6, 4), (6, 3), (7, 4), (7, 3), (8, 4)\}$ depends only on the pair (n, d) itself. Moreover, such a probability is exactly one of those listed in Table 4.

TABLE 4. Undetected error probabilities associated to non-trivial binary systematic AMDS codes.

(n,d)	$\mathrm{UEP}(p,C)$ of any $C \in \mathcal{C}_0(n,d)$
$(5,3)$	$p^5 - 3p^4 + 2p^3$
$(6,4)$	$3p^6 - 6p^5 + 3p^4$
$(6,3)$	$-p^6 + 6p^5 - 9p^4 + 4p^3$
$(7,4)$	$7(-p^7 + 3p^6 - 3p^5 + p^4)$
$(7,3)$	$p^7 - 7p^6 + 21p^5 - 21p^4 + 7p^3$
$(8,4)$	$15p^8 - 56p^7 + 84p^6 - 56p^5 + 14p^4$

Proof. The four sets $\mathcal{C}_0(5,3)$, $\mathcal{C}_0(6,4)$, $\mathcal{C}_0(6,3)$, $\mathcal{C}_0(7,3)$ contain only linear codes (see Lemma 26). By Remark 33 in this cases the undetected error probability is easily computed from the weight distribution given in Theorem 31. The value $\mathrm{UEP}(p,C)$ of a code $C \in \mathcal{C}_0(7,3)$ is computed by its distance distribution (Remark 33), given in the proof of Lemma 30. Since a code $C \in \mathcal{C}_0(8,4)$ is the parity-check code of a code in $\mathcal{C}(7,3)$ (Lemma 30) then the distance distribution of C is easily computed by Lemma 23:

$$\begin{array}{l|l|l} B_0(C)=1 & B_3(C)=0 & B_6(C)=0 \\ B_1(C)=0 & B_4(C)=14 & B_7(C)=0 \\ B_2(C)=0 & B_5(C)=0 & B_8(C)=1 \end{array}$$

This concludes the proof. □

9. CONCLUSIONS

By comparing Theorem 24 and Theorem 31 we note that the P- and the W- classification in fact coincide for the class binary systematic AMDS codes of interesting parameters, providing the same equivalence classes. In other words, the parameters $[n, 2^k, d]$ of any binary systematic AMDS code uniquely determine its weight distribution. For linear codes over $\mathbb{F}_q$ it is well-known (see [6], Theorem 9) that the weight distribution of a code C of parameters $[n, q^k, d]$ is completely determined by its weights $W_d(C),\ldots,W_{n-d^\perp}(C)$, where $d^\perp$ is the minimum distance of the dual code of C. In particular, if C is a binary linear AMDS code and its dual code $C^\perp$ is an AMDS code too, then the weight distribution of C is completely determined by its minimum weight $W_d(C)$. Theorem 31 in fact extends this result to the class of non-linear binary systematic AMDS codes without introducing any control over their dual codes.

ACKNOWLEDGMENT

The author would like to thank Emanuele Bellini, Simon Litsyn and Massimiliano Sala for useful suggestions and discussions.

REFERENCES

[1] T. L. Alderson, A. A. Bruen, *Maximal AMDS codes*. Applicable Algebra in Engineering, Communication and Computing, vol. 19, issue 2, pp. 87-98 (2008).
[2] R. D. Baker, J. H. van Lint, R. M. Wilson, *On the Preparata and Goethals codes*. IEEE Transactions on Information Theory, 29, 1983, pp. 342-345.
[3] E. Bellini, E. Guerrini, M. Sala, *A bound on the size of linear codes and systematic codes*. http://arxiv.org/abs/1206.6006.
[4] G Cohen, I. Honkala, S. Litsyn, A. Lobstein, *Covering Codes*. North-Holland Mathematical Library 54, 1997.
[5] S. Dodunekov, I. Landjev, *Near-MDS codes over some small fields*. Discrete Mathematics, vol. 213, pp. 55-65 (2000).
[6] A. Faldum, W. Willems, *Codes of Small Defect*. Designs, Codes and Cryptography, 10, pp. 341-350 (1997).
[7] E. Guerrini, M. Sala, *A classification of MDS binary systematic codes*. BCR preprint, UCC Cork, Ireland, 2006.
[8] W. C. Huffman, V. Pless, *Fundamentals of Error Correcting Codes*. Cambridge University Press, 2003.
[9] D. S. Kim, *Weight Distributions of Hamming Codes*. http://arxiv.org/abs/0710.1467.

[10] D. S. Kim, *Weight Distributions of Hamming Codes (II)*. `http://arxiv.org/abs/0710.1469`.

[11] F. J. MacWilliams, N. J. A. Sloane, *The Theory of Error-Correcting codes*. North Holland Mathematical Library, 1977.

[12] D. Nogin, M. Tsfasman, S. Vlăduţ, *Algebraic Geometry Codes, Basic Notions*. American Mathematical Society, Series *Mathematical Surveys and Monographs*, 2007.

[13] F. Preparata, *A class of optimum nonlinear double-error correcting codes*. Inform. Control 13, no. 13, pp. 378400 (1968).

[14] R. C. Singleton, *Maximum distance q-nary codes*. IEEE Transactions on Information Theory, 10, 1964, pp. 116-118.

[15] L. M. G. M. Tolhuizen, *On Maximum Distance Separable codes over alphabets of arbitrary size*. IEEE International Symposium on Information Theory, 1994.

[16] X-M. Wang, Y-X. Yang, *On the Undetected Error Probability of Nonlinear Binary Constant Weight Codes*. IEEE Transactions on Communications, vol. 42, n.7 (1994).